MATHEMATICS OF DATA SCIENCE

October 3, 2023

Program : Convex functions (Pt 2)

NO LECTURE ON OCTOBER 10

Next lecture : October 24

Tutorials : With me : Today + October 26 x2

With J. Lesca : Thursday + October 12 + October 26

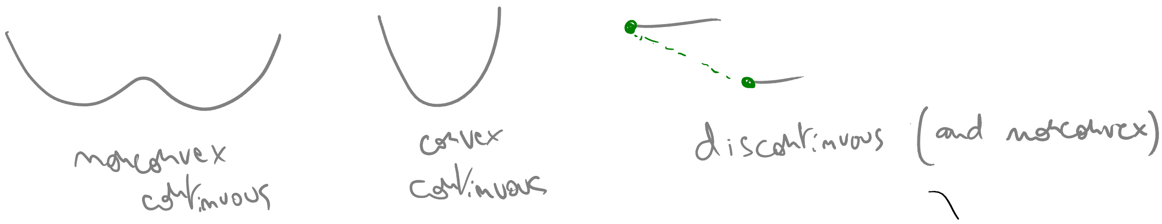
MORE ABOUT CONVEX FUNCTIONS

Previously
① Definition
② Extended-value functions

③ Convexity and regularity

Setting in this section: $X \subseteq \mathbb{R}^n$ convex set $\qquad f : \mathbb{R}^n \to \mathbb{R}$

Th) If f is convex on $\mathbb{R}^n$, then it is continuous.



Th) Suppose that f is differentiable in an open set containing X.

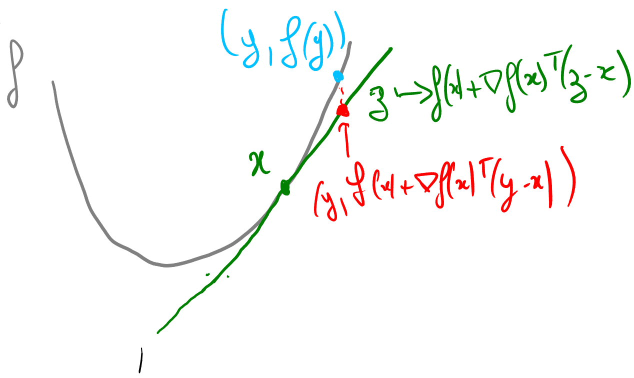
Then $[\, f$ is convex on $X \,]$

$$\iff \Big[\ \forall (x,y) \in X^2,\ f(y) \geq \underbrace{f(x) + \nabla f(x)^T (y-x)}\ \Big]$$

$u^T v = \sum_{i=1}^n u_i v_i$

"Gradient of f at x"

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_i}(x) \right]_{i=1}^{n} \in \mathbb{R}^n$$



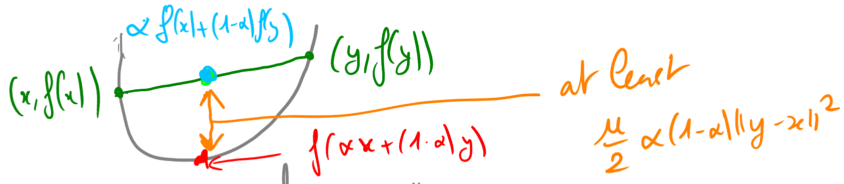
NB: When f is twice differentiable, there are other characterizations of convexity that involve second-order derivatives.

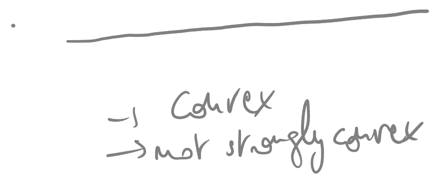
(4) Strong convexity

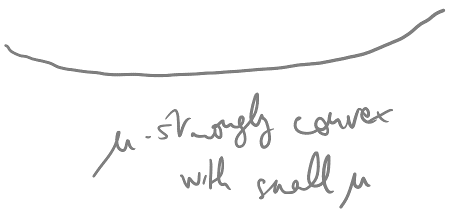
<u>Def</u>: Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ with X convex set. Let $\mu > 0$

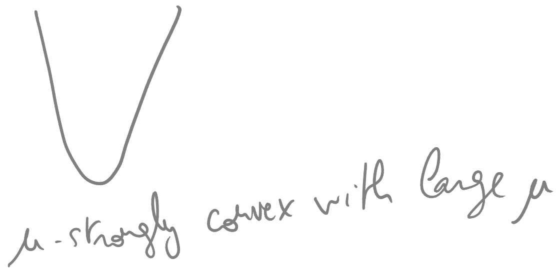
f is called μ-strongly convex if

$$\forall (x,y) \in X^2,\ \forall \alpha \in [0,1],\ \underbrace{f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)}_{\text{convexity}} - \frac{\mu}{2}\alpha(1-\alpha)\|y-x\|^2$$





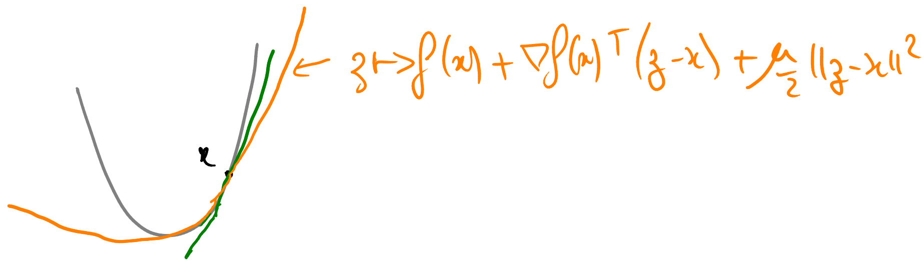




<u>Th</u>: If $f : X \to \mathbb{R}$ is differentiable on an open set containing X,

then $[f$ is μ-strongly convex on $X]$

$$\iff \left[\forall (x,y) \in X^2,\ f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2}\|y-x\|^2\right]$$



<u>Remarks</u>:

→ Linear functions are not strongly convex

→ Strongly convex functions form a "nicer" class of functions than convex functions

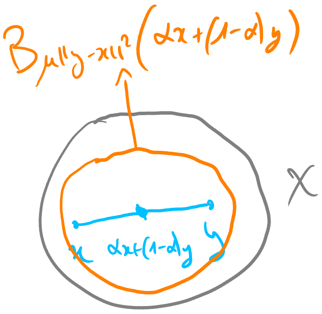
(but any strongly convex function is convex)

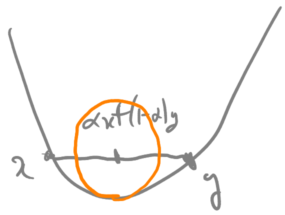
→ Thanks to the link between convex functions and convex sets, it is possible to define what a strongly convex set is.

Def: A set $X \subseteq \mathbb{R}^n$ is μ-strongly convex if ($\mu > 0$)

$$\forall (x,y) \in X^2,\ \forall \alpha \in [0,1],\ \forall z \in X \text{ such that } \|z\| \leq \mu \|y-x\|^2$$

$$\alpha x + (1-\alpha) y + z \in X$$





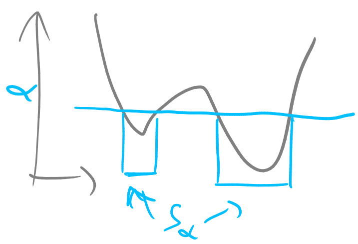
5 Level sets and quasi convexity

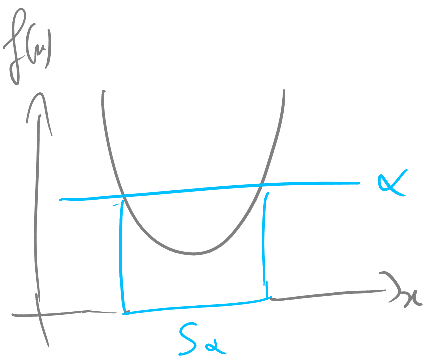
Def: Let $f : \mathbb{R}^n \to \bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

For any $\alpha \in \mathbb{R}$, the α-sublevel set of f, denoted by $S_\alpha(f)$, is given by

$$S_\alpha = S_\alpha(f) = \{ x \in \text{dom}(f) \mid f(x) \leq \alpha \}$$

$\text{dom}(f) = \{ x \mid f(x) < \infty \}$





Proposition: If $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex, then

$S_\alpha(f)$ is a convex set for any $\alpha \in \mathbb{R}$

⚠ Not an equivalence: there exist functions that have convex sublevel sets but that are not convex

Ex): $\mathbb{R}^2_{++} \to \mathbb{R}$
$x \mapsto -x_1 x_2$

Def: $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is called a quasiconvex function if $\forall \alpha \in \mathbb{R}$, $S_\alpha(f)$ is a convex set

NB: the definition can be restricted to a subset $X \subseteq \mathbb{R}^n$

Th: (characterizations of quasiconvexity)

- $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is quasiconvex (on dom f) if
$$\forall (x, y) \in \mathrm{dom} f^2, \forall \alpha \in [0, 1], \; f(\alpha x + (1-\alpha) y) \leq \max(f(x), f(y))$$

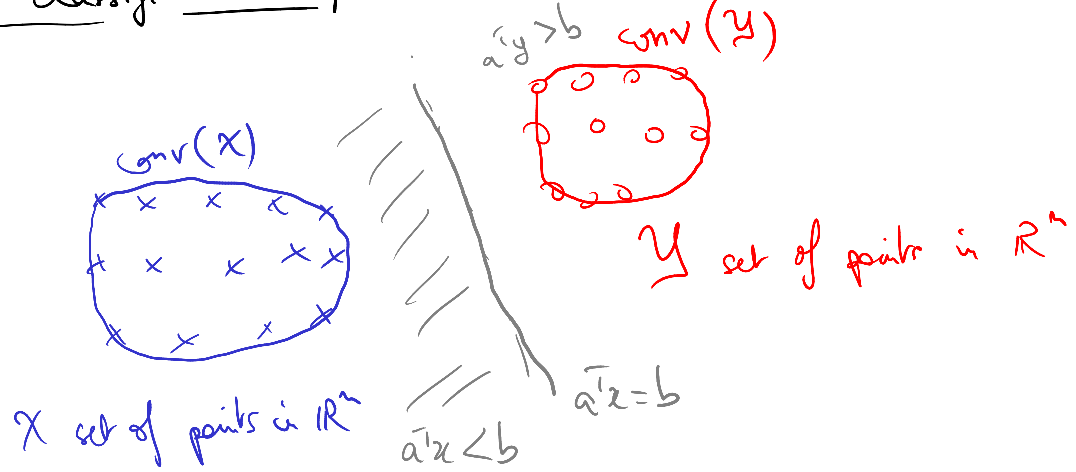
- $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ differentiable is quasiconvex if
$$\forall (x, y) \in \mathrm{dom} f^2, \quad f(y) \leq f(x) \implies \nabla f(x)^T (y - x) \leq 0$$

↳ The class of quasiconvex functions contains the class of convex functions (and the latter contains the class of strongly convex functions)

↳ there are many more subclasses and "superclasses" of convex functions

(6) One application of convexity

Basic linear classification setup



Goal: classify the points in $X \cup Y$ as belonging to either X or Y

Approach: Build a linear separator, that is a "line" that separates the points in X from those in Y

↳ In dimension $n \geq 1$, this concept of "line" is called a hyperplane and has the form

$$\{x \in \mathbb{R}^n \mid a^Tx = b\} \quad \text{for } a \in \mathbb{R}^n \text{ and } b \in \mathbb{R}$$

$\Rightarrow$ This is a convex set

↳ If conv(X) and conv(Y) are disjoint, then by the hyperplane separation theorem, there exist $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that $\{x \mid a^Tx = b\}$ separates X from Y, in the sense that

$$a^Tx < b \quad \text{if } x \in X$$
$$\text{and} \quad a^Ty > b \quad \text{if } y \in Y$$

↳ The best hyperplane is found using convex optimization

↳ Similar to what the last linear layer of certain neural architectures does.